*Материалы Российской НТК “Информатика и проблемы телекоммуникаций”, Новосибирск. 2008. – Т.1. – С. 84-87.*

# ПОСТРОЕНИЕ НАИБОЛЕЕ МОЩНОГО КРИТЕРИЯ ТИПА ХИ-КВАДРАТ ОТНОСИТЕЛЬНО ЗАДАННОЙ ПАРЫ АЛЬТЕРНАТИВ

Чаплина М.А., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: ai-no-kasumi@yandex.ru

Целью работы явилось исследование мощности критериев типа $\chi^2$ при группировании, основанном на условной максимизации параметра нецентральности, сравнение с мощностью критериев при асимптотически оптимальном (АОГ) и равновероятном (РВГ) группировании для простых и сложных проверяемых гипотез в зависимости от числа интервалов группирования. Было проанализировано, каким образом можно максимизировать мощность критериев типа $\chi^2$ при заданной паре близких конкурирующих гипотез $H_0$ и $H_1$.

При использовании критериев согласия типа $\chi^2$ область определения случайной величины разбивается на $k$ интервалов граничными точками $x_0 < x_1 < \ldots < x_{k-1} < x_k$. Статистика $X_n^2$ Пирсона вычисляется в соответствии с соотношением:

$$X_n^2 = n\sum_{i=1}^{k}\frac{(n_i / n - P_i(\theta))^2}{P_i(\theta)}, \tag{1}$$

где $n_i$ – количество наблюдений, попавших в интервал, $P_i$ - вероятность попадания наблюдения в $i$-й интервал, $n=\sum_{i=1}^{k} n_i$, $\sum_{i=1}^{k} P_i(\theta)=1$.

При справедливой простой гипотезе $H_0$ предельное распределение статистики $G(X_n^2 | H_0)$ есть $\chi_r^2$-распределение с числом степеней свободы $r=k-1$. Если по выборке оценивалось $m$ параметров закона в результате минимизации статистики $X_n^2$, статистика подчиняется $\chi_r^2$-распределению с $r=k-m-1$ степенями свободы. При справедливой конкурирующей гипотезе $H_1$ предельное распределение $G(X_n^2 | H_1)$ представляет собой нецентральное $\chi_r^2$-распределение с тем же числом степеней свободы и параметром нецентральности:

$$\lambda = n\sum_{i=1}^{k}\frac{(P_i^1(\theta_1) - P_i(\theta))^2}{P_i(\theta)}, \tag{2}$$

где $P_i^1(\theta_1)$ – вероятность попадания в интервал при справедливой гипотезе $H_1$.

Никулиным [2] предложено такое видоизменение стандартной статистики (1), при котором предельное распределение есть обычное распределение $\chi^2_{k-1}$. Неизвестные параметры распределения $F(x,\theta)$ в этом случае должны оцениваться по негруппированным данным методом максимального правдоподобия. Предложенная статистика отличается от $X_n^2$ только при сложных гипотезах и имеет вид [2,3]:

$$Y_n^2 = X_n^2 + n^{-1} a^{\mathrm{T}}(\theta)\mathbf{\Lambda}(\theta)a(\theta). \qquad (3)$$

При справедливой альтернативе $H_0$ предельное распределение этой статистики - нецентральное $\chi^2_{k-1}$-распределение с параметром нецентральности [2]:

$$\lambda = \sum_{j=1}^{r} \frac{c_j^2(\theta)}{p_j} + d^T(\theta)G^{-1}(\theta)d(\theta), \quad d_j(\theta) = \sum_{i=1}^{r} \frac{\omega_{ij}(\theta)c_i(\theta)}{p_i}. \qquad (4)$$

Мощность критериев $1-\beta$, где $\beta$ - вероятность ошибки второго рода, зависит от проверяемой гипотезы, объема выборки $n$, числа интервалов группирования $k$ и от способа группирования. С позиции обеспечения максимальной мощности критерия сравнивались различные способы группирования: асимптотически оптимальное (АОГ), равновероятное (РВГ) и группирование, обеспечивающее максимум параметра нецентральности. Первые два опираются на знание закона распределения, соответствующего $H_0$, третий – еще и на знание закона распределения, соответствующего $H_1$.

Мощность критерия является неубывающей функцией параметра нецентральности [1], поэтому задачу максимизации мощности можно свести к задаче условной максимизации параметра нецентральности:

$$\max_{x_{(0)} \le x_{(1)} \le .. \le x_{(k)}} \lambda = \min_{x_{(0)} \le x_{(1)} \le .. \le x_{(k)}} \left( -n \sum_{i=1}^{k} \frac{(P_i^1(\theta_1) - P_i(\theta))^2}{P_i(\theta)} \right). \qquad (5)$$

В случае использования критерия Никулина задача максимизации имеет вид:

$$\max_{x_{(0)} \le x_{(1)} \le .. \le x_{(k)}} \lambda = \min_{x_{(0)} \le x_{(1)} \le .. \le x_{(k)}} \left( \sum_{j=1}^{r} \frac{c_j^2(\theta)}{p_j} + d^T(\theta)G^{-1}(\theta)d(\theta) \right). \qquad (6)$$

Для решения этих задач использовался метод вращающихся координат, для учета ограничений - метод штрафных функций. Результат решения зависит от соответствующих распределений, объема выборки, числа интервалов и начального приближения граничных точек. Сложность задачи определяется ее многоэкстремальностью. Для каждой пары конкурирующих гипотез существует оптимальное число интервалов (оно зависит от объема выборки), поэтому для обеспечения максимальной мощности критерия эту задачу необходимо решать для различных $k$.

На основе проведенных исследований показано, что группирование, обеспечивающее максимум параметра нецентральности, позволяет достичь наибольшей мощности, причем в случае сложной гипотезы более высокой мощностью обладает критерий Никулина.

В приводимой таблице представлены значения мощности при проверке простой и сложной гипотез, когда проверяемой гипотезе соответствует нормальное распределение, а конкурирующей - логистическое [3]:

Таблица 1. Мощность критериев Пирсона (1) и Никулина (3) при различных способах группирования

| $\alpha$ | *Простая гипотеза* | | | *Сложная гипотеза* | | |
|---|---|---|---|---|---|---|
| | *n*=100 | *n*=200 | *n*=1000 | *n*=200 | *n*=500 | *n*=1000 |
| *Мощность критерия Пирсона при k*=9 и АОГ | | | | | | |
| 0.15 | 0.269 | 0.670 | 0.917 | 0.404 | 0.516 | 0.930 |
| 0.1 | 0.204 | 0.589 | 0.880 | 0.323 | 0.431 | 0.898 |
| 0.05 | 0.129 | 0.464 | 0.806 | 0.219 | 0.314 | 0.834 |
| 0.025 | 0.084 | 0.359 | 0.723 | 0.146 | 0.225 | 0.760 |
| 0.01 | 0.050 | 0.249 | 0.608 | 0.085 | 0.141 | 0.654 |
| *Мощность критерия Пирсона при k*=9 и РВГ | | | | | | |
| 0.15 | 0.210 | 0.483 | 0.747 | 0.309 | 0.541 | 0.828 |
| 0.1 | 0.152 | 0.392 | 0.673 | 0.232 | 0.445 | 0.757 |
| 0.05 | 0.083 | 0.273 | 0.547 | 0.136 | 0.309 | 0.629 |
| 0.025 | 0.046 | 0.186 | 0.435 | 0.079 | 0.208 | 0.502 |
| 0.01 | 0.020 | 0.109 | 0.310 | 0.039 | 0.120 | 0.358 |
| *Мощность критерия Пирсона при k*=9 и решении задачи (5) | | | | | | |
| 0.15 | 0.289 | 0.687 | 0.925 | 0.421 | 0.723 | 0.973 |
| 0.1 | 0.215 | 0.592 | 0.892 | 0.338 | 0.638 | 0.902 |
| 0.05 | 0.157 | 0.505 | 0.823 | 0.227 | 0.505 | 0.841 |
| 0.025 | 0.102 | 0.367 | 0.751 | 0.175 | 0.402 | 0.775 |
| 0.01 | 0.073 | 0.263 | 0.622 | 0.092 | 0.295 | 0.670 |
| *Мощность критерия Никулина при k*=9 и решении задачи (6) | | | | | | |
| 0.15 | – | – | – | 0.594 | 0.836 | 0.986 |
| 0.1 | – | – | – | 0.535 | 0.794 | 0.955 |
| 0.05 | – | – | – | 0.452 | 0.727 | 0.931 |
| 0.025 | – | – | – | 0.368 | 0.655 | 0.901 |
| 0.01 | – | – | – | 0.282 | 0.562 | 0.852 |

Литература


1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
2. Мирвалиев М., Никулин М.С. Критерии согласия типа хи-квадрат / Заводская лаборатория. 1992. Т. 58. № 3. – С.52-58.
3. Лемешко Б.Ю., Лемешко С.Б., Постовалов С.Н. Мощность критериев согласия при близких альтернативах // Измерительная техника. 2007. № 2. – С.22-27.